

2019年5月30日

分子モデリングと分子シミュレーション

分子動力学法実習

東京大学大学院農学生命科学研究科

アグリバイオインフォマティクス

教育研究プログラム

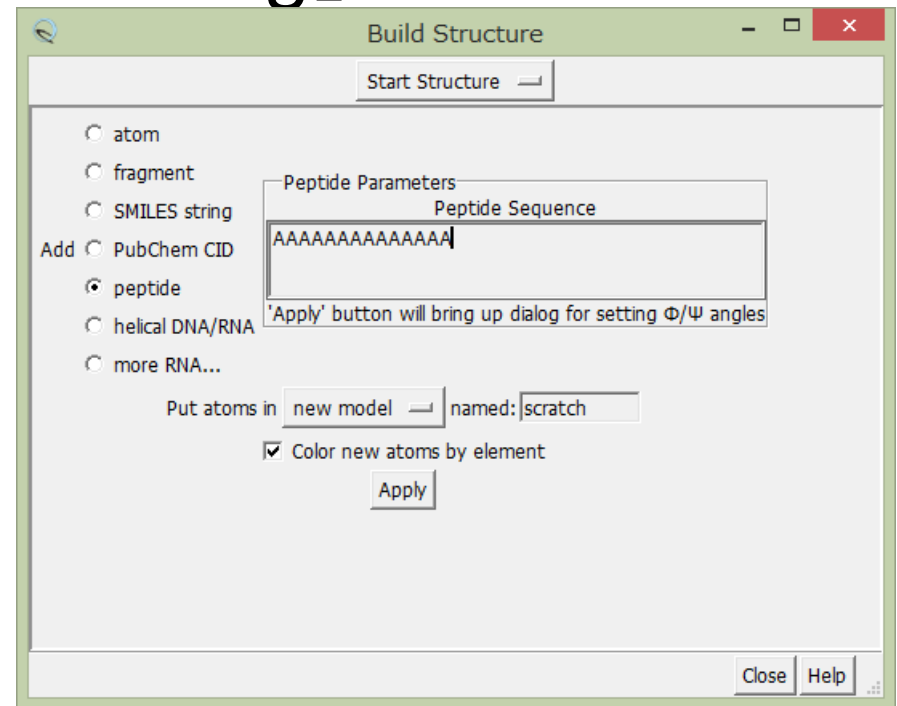
寺田 透

本日の講義内容

- ペプチドの分子動力学シミュレーション
- 水溶液環境のモデル
- ペプチドの分子動力学シミュレーション
 - 課題1
- タンパク質の分子動力学シミュレーション
 - 課題2
- シミュレーションの高速化
- シミュレーション実行上の注意点

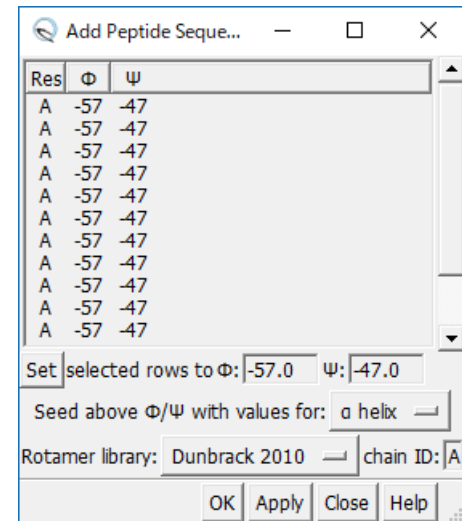
ペプチドの生成(1)

1. UCSF Chimeraを起動
2. 「Tools」→「Structure Editing」→「Build Structure」を選択
3. 「Start Structure」、
「peptide」を選択し、Peptide Sequenceに「A」を14個入力し「Apply」



ペプチドの生成(2)

4. Add Peptide Sequenceウィンドウで主鎖の二面角を指定し「OK」(ここでは α helix構造を作るのでデフォルトのままが良い)
5. メイン画面に生成されたペプチドの構造が現れるので、「Actions」→「Atoms/Bonds」→「show」、
「Actions」→「Ribbon」→「hide」
6. 「File」→「Save PDB」でala14.pdbとしてデスクトップに保存



力場パラメータの割り当て

1. 「Tools」→「Structure Editing」→「AddH」を選択し「OK」→水素原子を付加
2. 「Tools」→「Structure Editing」→「Add Charge」を選択し、Standard residuesの力場に「AMBER ff14SB」を指定し、「OK」
3. 「Tools」→「Amber」→「Write Prmtop」を選択し、Folderが「C:¥Users¥student¥Desktop」になっていることを確認し、File nameに「ala14」、Select force field typeに「AMBER ff14SB」を指定し「Save」

分子動力学シミュレーションの手順

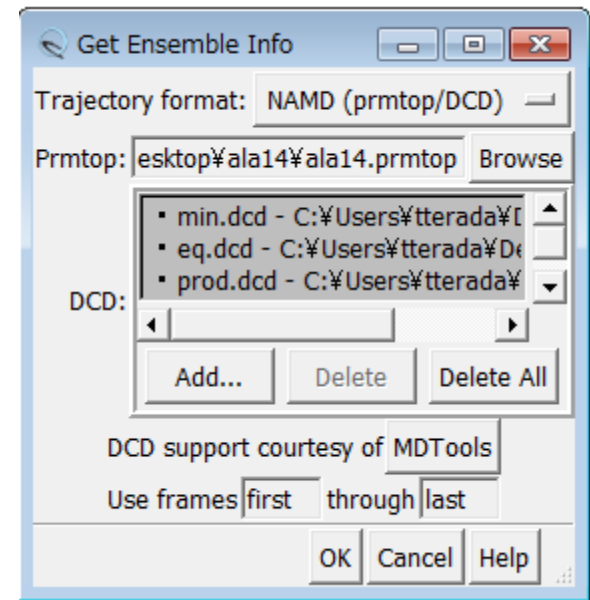
1. 講義のページからnamd2.exeとtcl85t.dllをダウンロードし、デスクトップに保存
2. 講義のページからala14.zipをダウンロードし、デスクトップに保存
3. ala14.zipをダブルクリックして解凍し、生成されたフォルダ(ala14)に、先に作成した、ala14.prmtopとala14.inpcrdを移動
4. run.batをダブルクリックして実行

参考：ソフトウェア

- NAMD
 - 本講義で使用。無料。
 - AMBER、CHARMM力場に対応
 - <http://www.ks.uiuc.edu/Research/namd/>
- Gromacs
 - 無料
 - AMBER、CHARMM、GROMOS力場に対応
 - <http://www.gromacs.org/>
- 他にAMBER、CHARMM、Desmondなど

シミュレーション結果の表示(1)

1. Chimeraを起動する(起動済みの場合は、「File」→「Close Session」を選択)
2. 「Tools」→「MD/Ensemble Analysis」→「MD Movie」を選択
3. Trajectory formatに「NAMD (prmtop/DCD)」、Prmtopにala14.prmtopを、DCDにmin.dcd、eq.dcd、prod.dcdをこの順で指定し、「OK」



シミュレーション結果の表示(2)

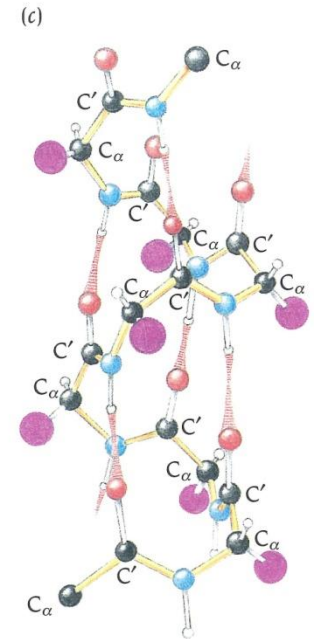
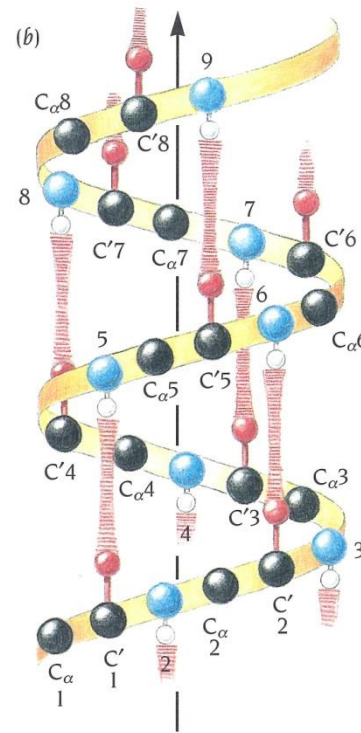
4. MD Movieウィンドウにある再生ボタンをクリックし、最初はリボンモデルのまま、運動の様子を観察せよ
5. 「Actions」→「Atoms/Bonds」→「show」、
「Actions」→「Ribbon」→「hide」でスティックモデルに変更せよ
6. 運動の様子を観察せよ

初期構造からのずれ (RMSD)

1. 「Select」→「Atom Specifier」を選択し、Atom Specifier to Selectに「@CA」と入力し「OK」→Ca原子を選択
2. MD Movieのメニューの「Analysis」→「Plot」→「RMSD」を選択
3. MD Plotsウィンドウにある「Plot」ボタンをクリック

参考：水素結合距離の測定

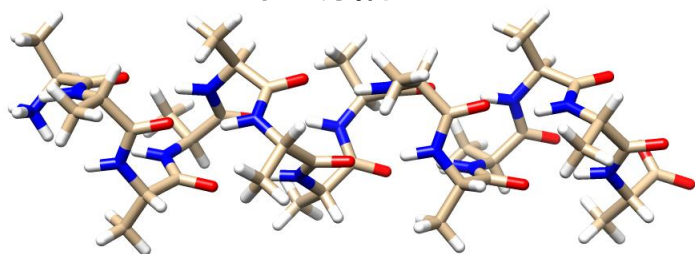
1. MD Movieウインドウの「Analysis」→「Plot」→「Distances」を選択
2. 原子間距離を測りたい原子のペアの一方をCtrlキーを押しながら左クリックで選択し、もう一方をCtrlキーとShiftキーを押しながら左クリックで選択
 - distance.comで水素結合原子ペアを表示すると良い
3. MD Plotsウインドウにある「Plot」ボタンをクリック



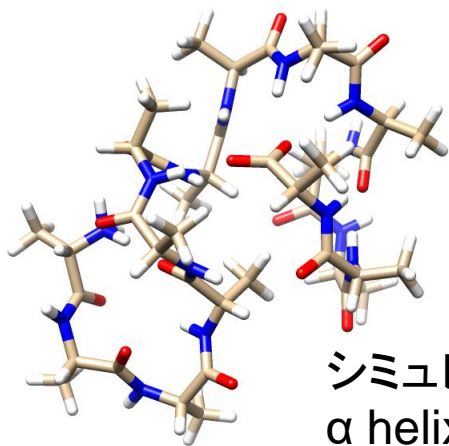
α helixでは*i*番目のカルボニル酸素と*i*+4番目のアミド窒素が水素結合を形成する

シミュレーションの結果

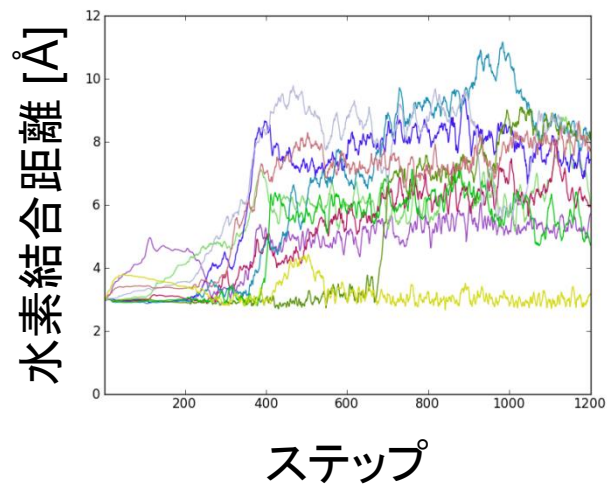
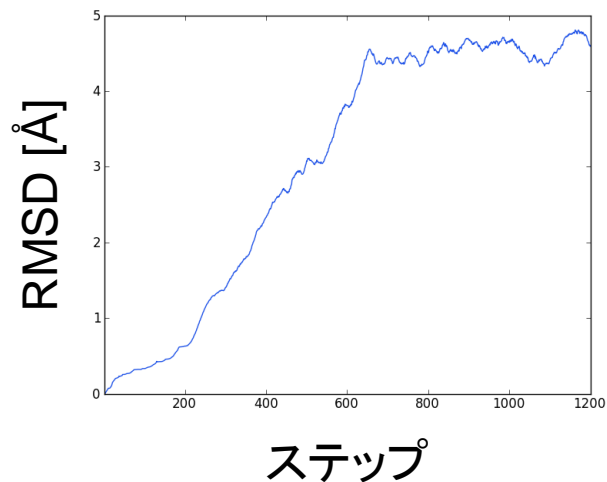
初期構造



最終構造



シミュレーションによって
 α helix構造が壊れていることに注意



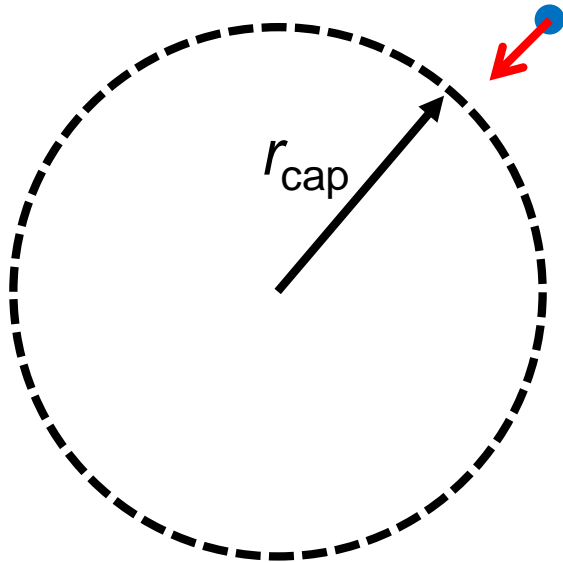
水溶液環境のモデル(1)

- 今回のシミュレーションは真空中で行われており、水分子による溶媒効果は考慮されていない
- 生体分子のシミュレーションにおいては、水溶液環境を適切なモデルを用いて再現する必要がある

水溶液環境のモデル(2)

- 現在以下の方法がよく用いられている
- 水分子を陽に配置
 - 球状に配置
 - 直方体状に配置→周期境界条件
- 溶媒和自由エネルギーを近似的に求める
 - 非極性項→溶媒接触表面積に比例
 - 極性項→連続誘電体モデル
 - Poisson-Boltzmann方程式
 - Generalized Bornモデル

球状の配置



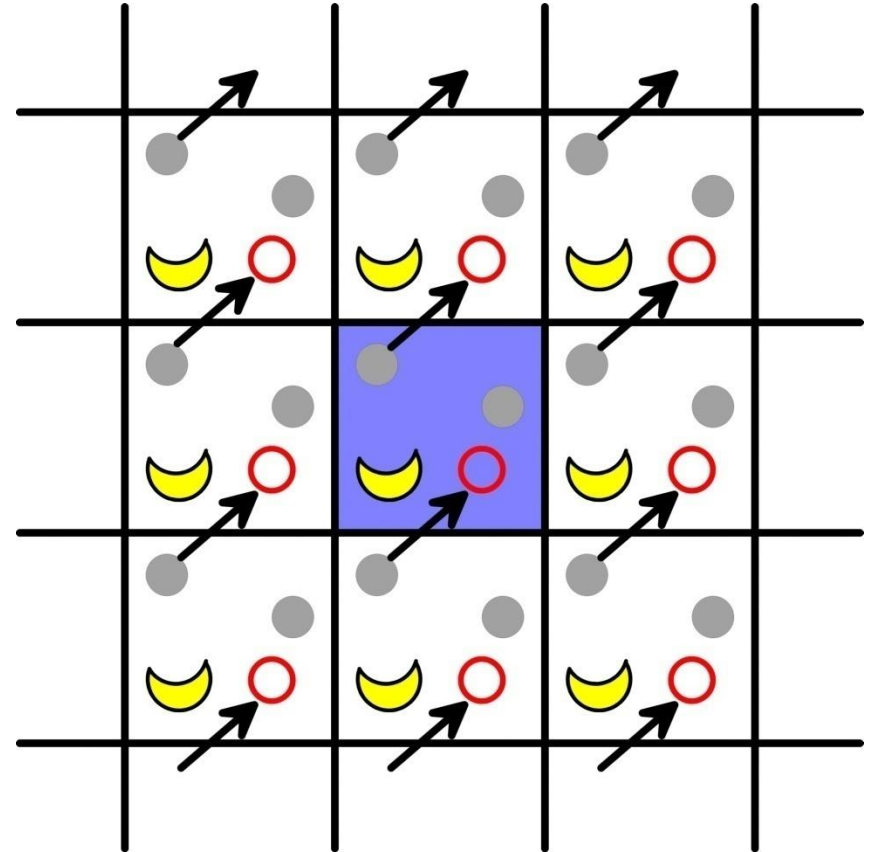
$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

$$E_{\text{cap}} = \begin{cases} 0 & r < r_{\text{cap}} \\ k(r - r_{\text{cap}})^2 & r \geq r_{\text{cap}} \end{cases}$$

- 水分子の“蒸発”を防ぐため、分子が半径 r_{cap} の球の外側に出て行こうとすると、系の中心に向けて束縛力をかける
- 系の表面に位置する水分子は中心付近の水とは異なる環境に置かれる

周期境界条件

- 中央のセルと同じものが無限に繰り返す
- セルから出て行った分子は、そのセルの反対側から入る
- どの分子も同じ環境
- 系が隣接セルからの影響を感じないように、系のサイズを十分に大きくする必要があります



圧力の計算

$$F = E - TS, \quad dE = -PdV + TdS$$

$$dF = dE - TdS - SdT = -PdV - SdT$$

$$P = -\left(\frac{\partial F}{\partial V}\right)_T = k_B T \left(\frac{\partial \ln Z}{\partial V}\right)_T = \frac{k_B T}{Z} \left(\frac{\partial Z}{\partial V}\right)_T$$

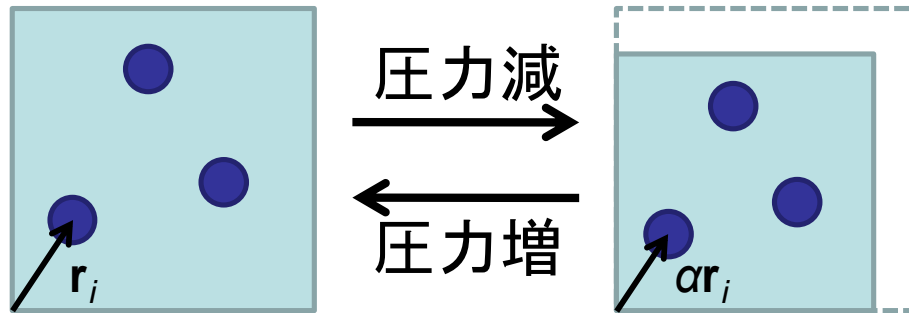
$$= \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{f}_i \quad \text{ビリアル定理}$$

$$= \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i=1}^N \sum_{j=i+1}^N \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \quad \text{周期境界条件ではこちらを使う}$$

相互作用のない系(理想気体)では、 $PV = Nk_B T = nRT$

圧力の制御

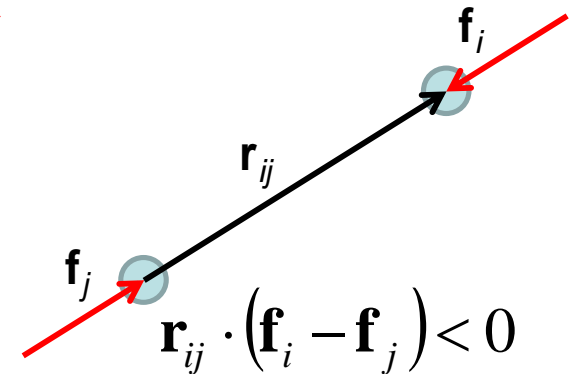
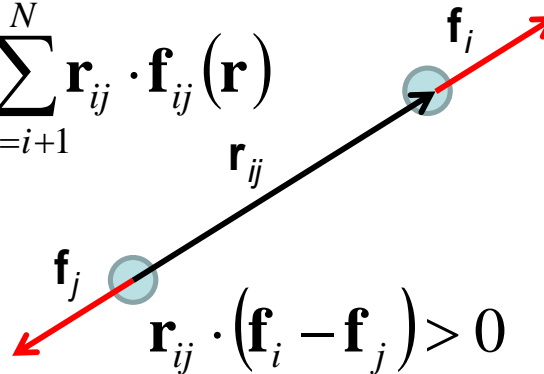
- 周期境界条件における、セルの大きさを変化させることで圧力を制御する



分子の重心位置も同様にスケールされる
分子内の原子の相対位置は変化しない

- 瞬間的に $P < 0$ となることがある

$$P = \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i=1}^N \sum_{j=i+1}^N \mathbf{r}_{ij} \cdot \mathbf{f}_{ij}(\mathbf{r})$$



水溶液中のシミュレーション(1)

1. Chimeraを起動し、ala14.pdbを開く
2. Stick表示に変更し、水素原子を付加する
3. 「Tools」→「Structure Editing」→「Solvate」を選択し、Solvate methodに「Box」、Solvent Modelに「TIP3PBOX」、Box sizeに「6」を入力し、「OK」
4. 「Tools」→「Structure Editing」→「Add Charge」で、Standard residuesに「AMBER ff14SB」を指定し、「OK」

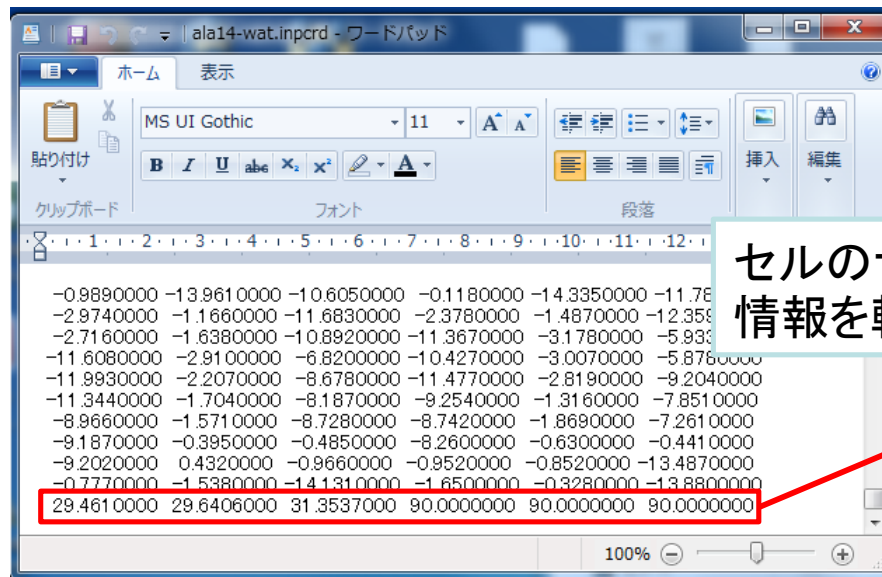
水溶液中のシミュレーション(2)

5. 「Tools」→「Amber」→「Write Prmtop」を選択し、Folderが「C:¥Users¥student¥Desktop」になっていることを確認し、File nameに「ala14-wat」、Select force field typeに「AMBER ff14SB」を指定し「Save」
6. 「File」→「Save PDB」を選択し、デスクトップにala14-wat.pdbとして保存
7. 講義のページから、ala14-wat.zipをダウンロードし、デスクトップに保存し、解凍
8. ala14-wat.prmtop、ala14-wat.inpcrd、ala14-wat.pdbを生成したala14-watフォルダに移動

水溶液中のシミュレーション(3)

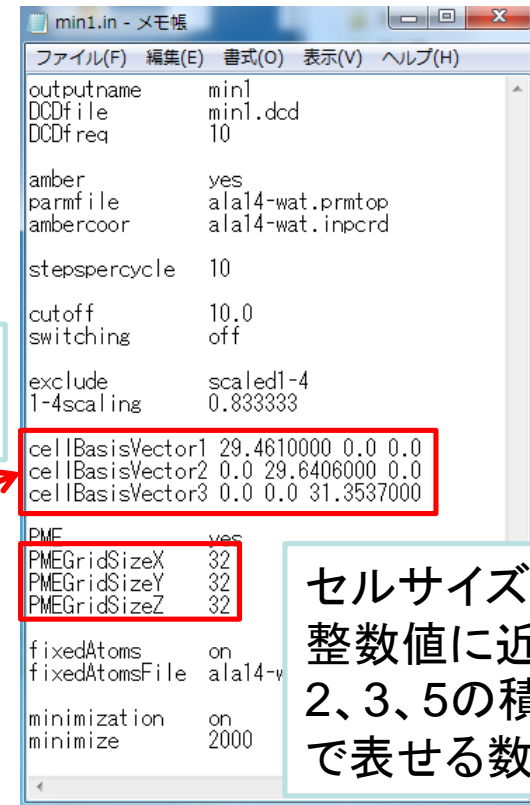
9. ala14-watフォルダを開き、restraint.plをダブルクリック→ala14-wat_rest.pdbが生成

10. min1.inを以下の通り修正



セルのサイズの
情報を転記

ala14-wat.inpcrdをワードパッド
で開き末尾を表示



セルサイズの
整数値に近く
2、3、5の積
で表せる数

水溶液中のシミュレーション(4)

11. run.batをダブルクリックし、シミュレーションを以下の順に実行(約3分)

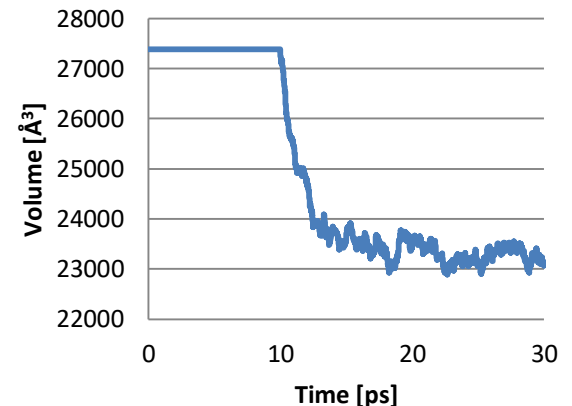
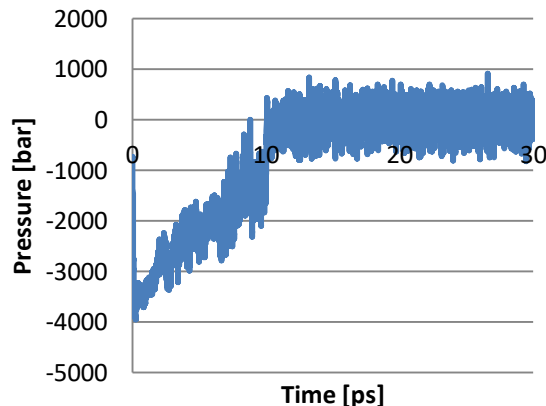
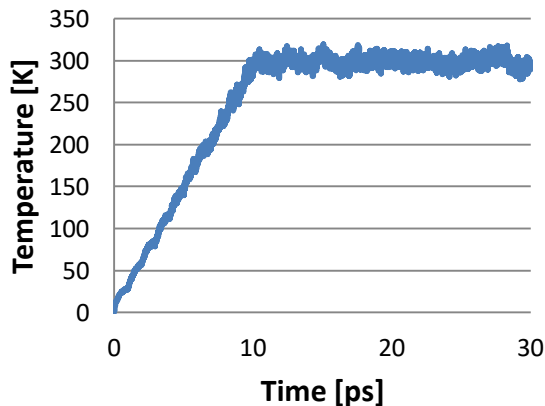
- ① エネルギー最小化(水分子のみ)(min1)
- ② エネルギー最小化(全体)(min2)
- ③ 平衡化(0→300 K)(eq1, 10 ps)
- ④ 平衡化(定圧)(eq2, 10 ps)
- ⑤ プロダクション(prod, 10 ps)

結果の解析

1. コマンドプロンプトを起動し、以下を実行

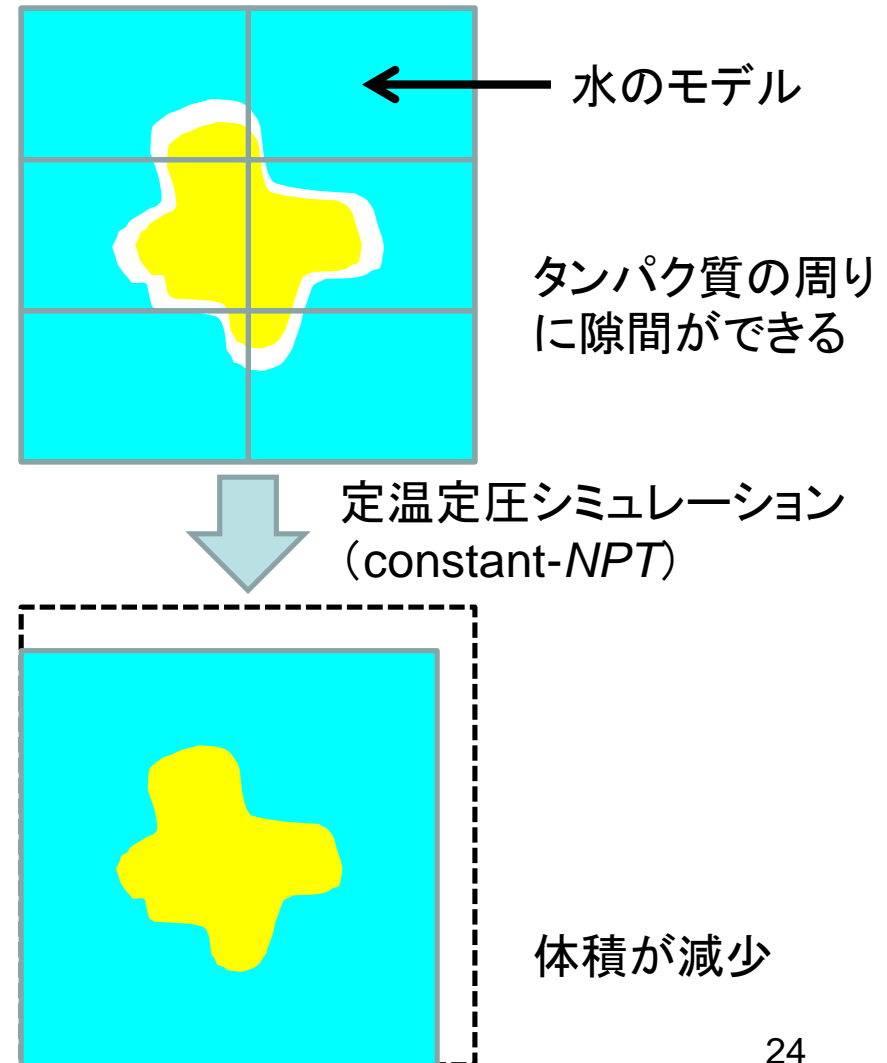
```
> cd Desktop¥ala14-wat  
> energy.pl eq1.log eq2.log prod.log
```

2. energy.csvが生成されるので、Excelで開く



平衡化における体積の変化

- 水を配置する際、少数の水分子を小さな系で平衡化したモデルタンパク質の周囲にあてはめているが、タンパク質の原子と衝突する水分子は機械的に取り除いているため、配置した水分子とタンパク質の間に隙間ができる
- 定温定圧シミュレーションを行い、水分子の配置を最適化すると隙間が埋まり、体積が減少する



課題1

- 平衡化 (eq1、eq2) とプロダクション (prod) における、温度 (TEMP) と圧力 (PRESSURE)、体積 (VOLUME) の時間変化をプロットせよ
 - 時間刻み Δt は 2 fs
- エネルギー最小化 (min1、min2)、平衡化 (eq1、eq2)、プロダクション (prod) における、Ca RMSD の変化をプロットせよ
- これらのプロットから何が言えるか考察せよ

計算時間(1)

- 対象: 球状に配置した水分子(TIP3Pモデル)
- Amber 18のSanderモジュール使用
- 計算にはIntel Xeon Processor 6コア使用
- 時間刻み Δt は0.5 fs
- 1 psの計算にかかる時間(単位は秒)を計測

原子数	T_{total} [s]	比率	T_{nb} [s]	$T_{\text{nb}}/T_{\text{total}}$
3087	30.3	1.0	30.0	0.996
6066	119.0	3.9	118.4	0.998
10608	360.9*	11.9	359.8	0.999

*1 nsあたり4.2日かかる

分子シミュレーションの効率化

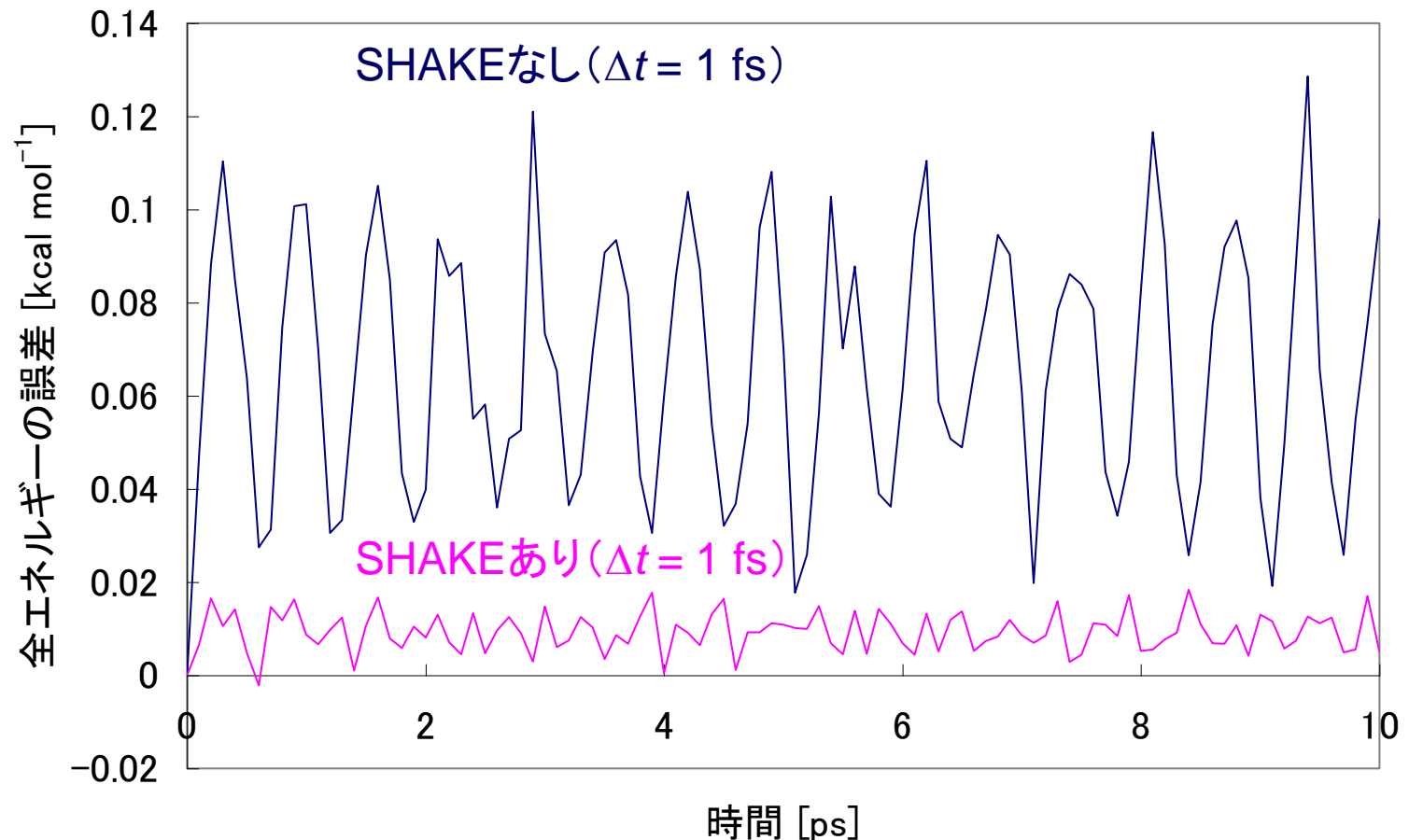
- 時間刻み Δt を長くする
 - SHAKE法
 - 多重時間積分法
- 非共有結合相互作用の計算の近似
 - カットオフ法
 - 多重極子展開法
 - Particle mesh Ewald (PME) 法
- 本講義では赤枠の3つの方法について解説

SHAKE法

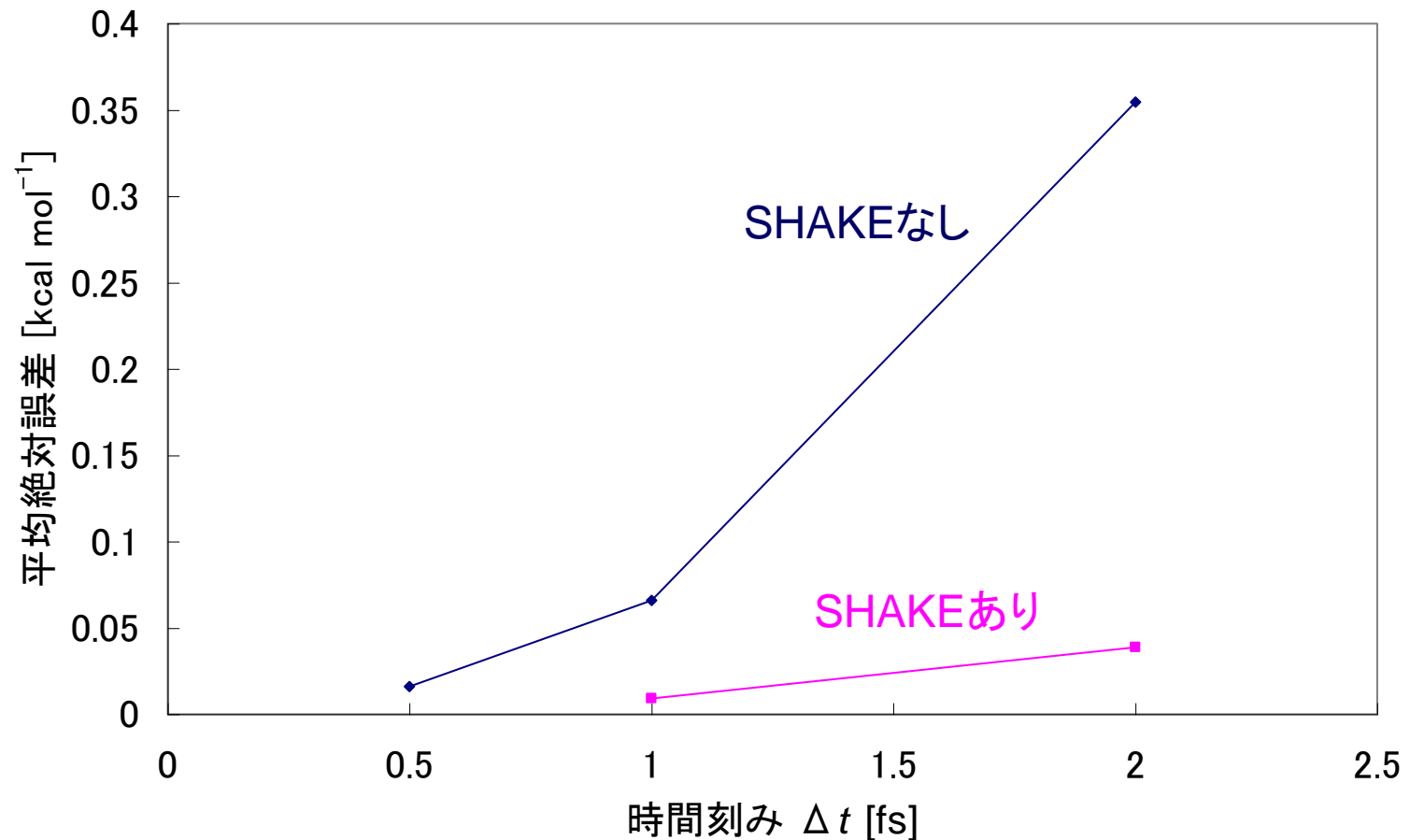
- 時間刻みは、最も速い運動の周期の10分の1から20分の1
- 最も速い運動は、X-H伸縮運動
→周期は約10 fs → $\Delta t = 0.5 \sim 1$ fs
- 次に速い運動は、X-X伸縮運動
→周期は約20 fs
- SHAKE法によりX-H結合長を固定
→長い時間刻み ($\Delta t = 2$ fs) の使用が可能

SHAKEの適用例

Methanolの分子動力学シミュレーション (constant- NVE)
における全エネルギーの誤差 (初期値との差) の推移



SHAKEの適用例



SHAKEを用いると時間刻み2 fsでもSHAKEなしの0.5 fsに匹敵する精度が得られる

NAMDにおける設定(1)

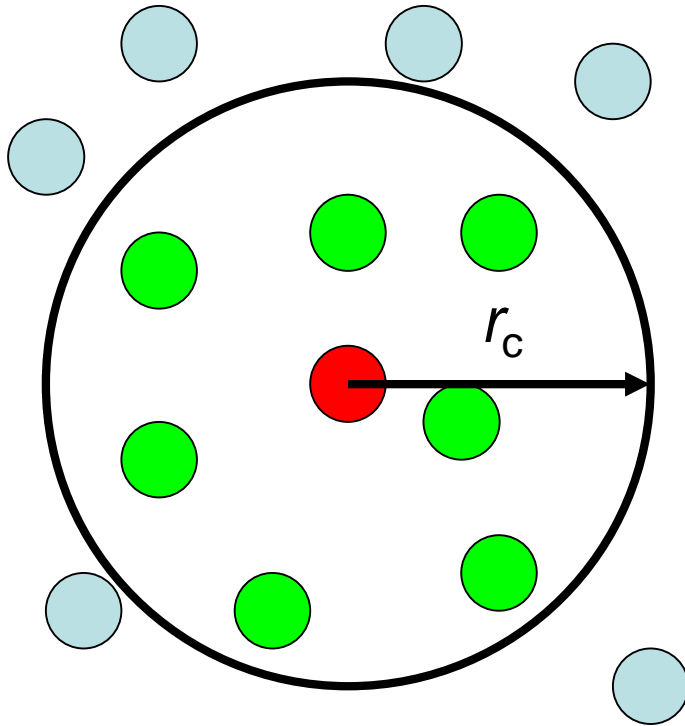
- SHAKEを使う場合は以下の設定を行う

```
rigidBonds      all  
useGroupPressure  yes
```

非共有結合相互作用の扱い

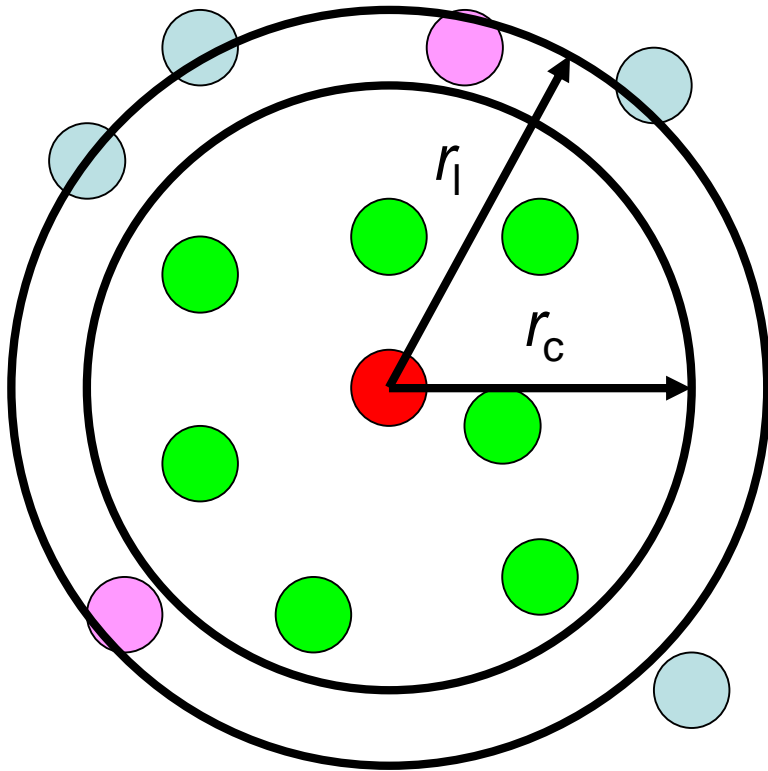
- 非共有結合相互作用は、原子のペアについて計算する必要がある
→ N 原子系では $N(N-1)/2$ のペア
- 非共有結合相互作用は距離が離れるほど弱くなる (van der Waals 引力は r^{-6} に比例、静電相互作用は r^{-1} に比例)
- 離れている原子同士は相互作用しないとみなす → カットオフ法

カットオフ法



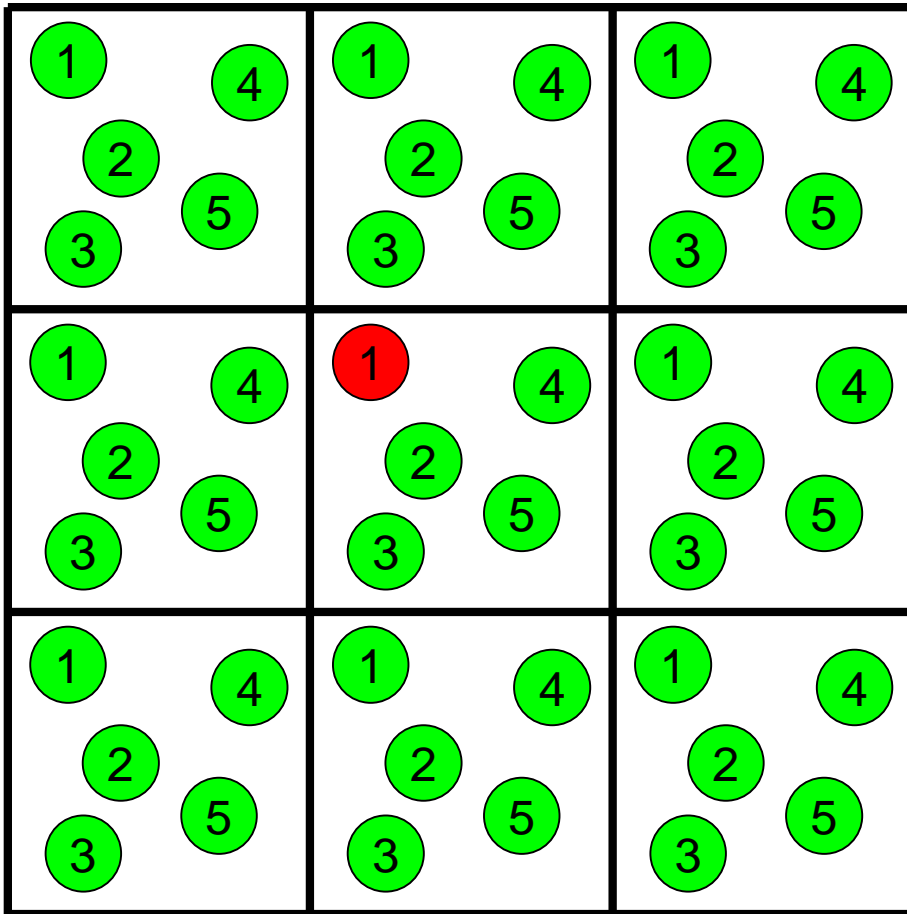
- 原子から半径 r_c の範囲内にある原子との非共有結合相互作用の計算を行う
- この範囲にある原子の平均個数を M とすると、非共有結合相互作用の計算量は $N(N-1)/2$ から NM に減少する

ペアリストの作成



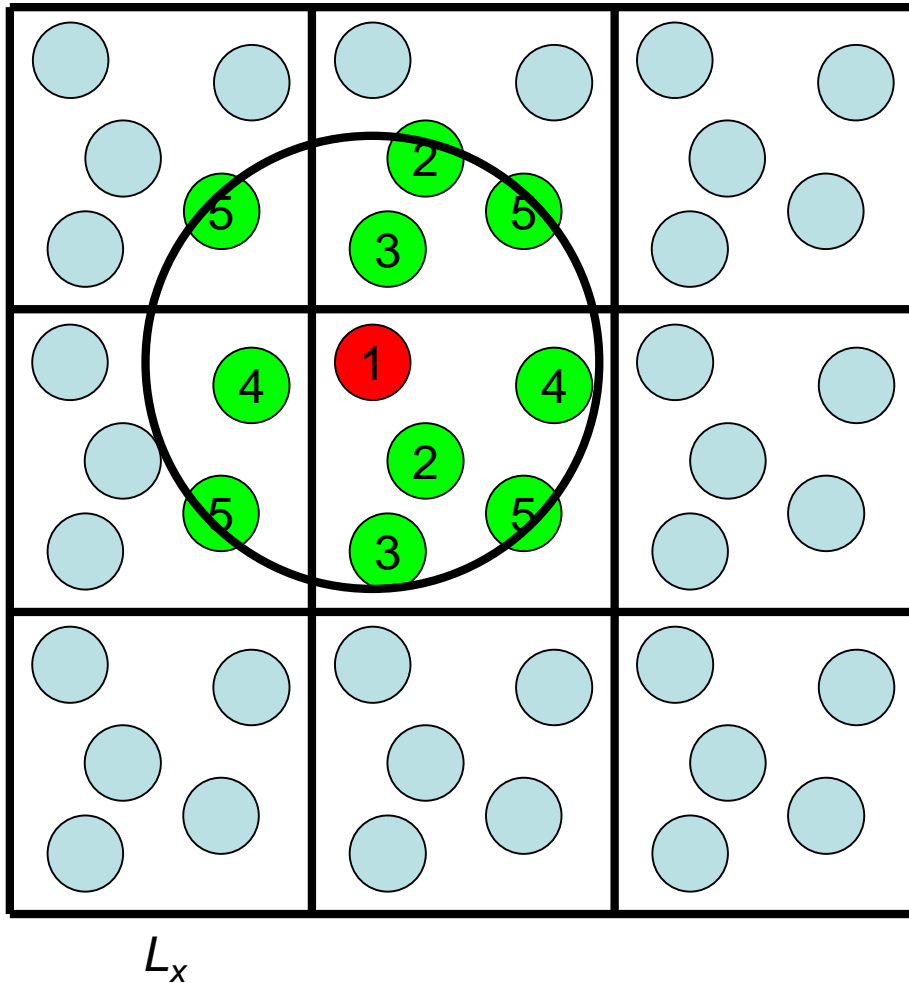
- カットオフ半径 r_c 以内にある原子ペアのリストを作成する必要がある
- この計算量は $N(N-1)/2$
- カットオフ半径 r_c より外側の半径 r_l の範囲でリストを作っておき、原子の最大移動度が $r_l - r_c$ を超えた時にリストを更新するようにすると計算量を削減できる

周期境界条件の場合(1)



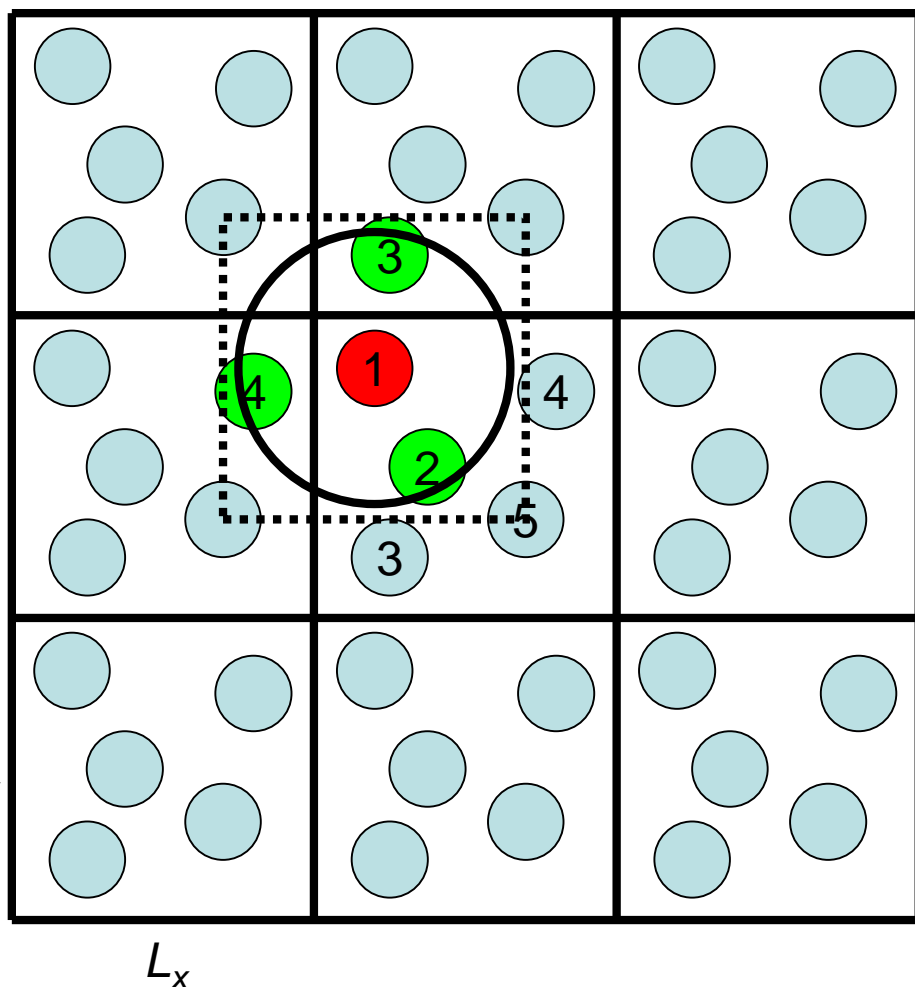
- 周期境界条件では基本セルのコピーが無限に続くので全ての原子ペアについて相互作用を**近似せずに**直接計算することは不可能

カットオフ法の適用



- カットオフ半径によっては、基本セルの周辺イメージセルも考慮する必要がある
(左の例では $26N^2 + N(N-1)/2$ ペアの計算が必要)

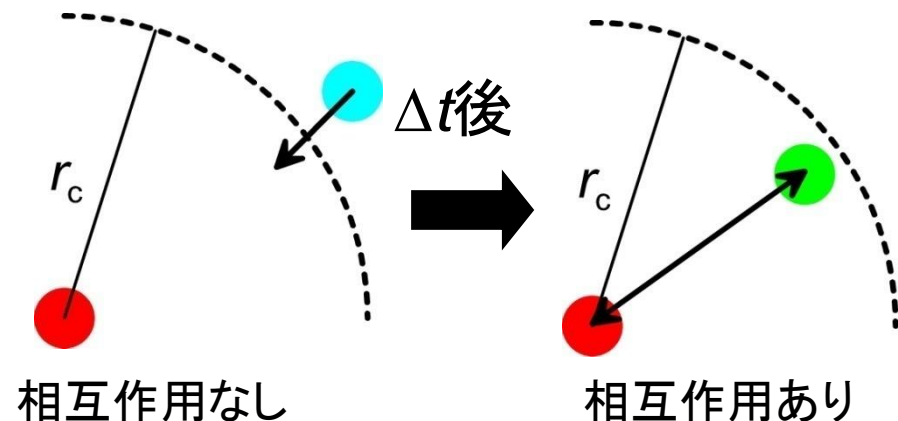
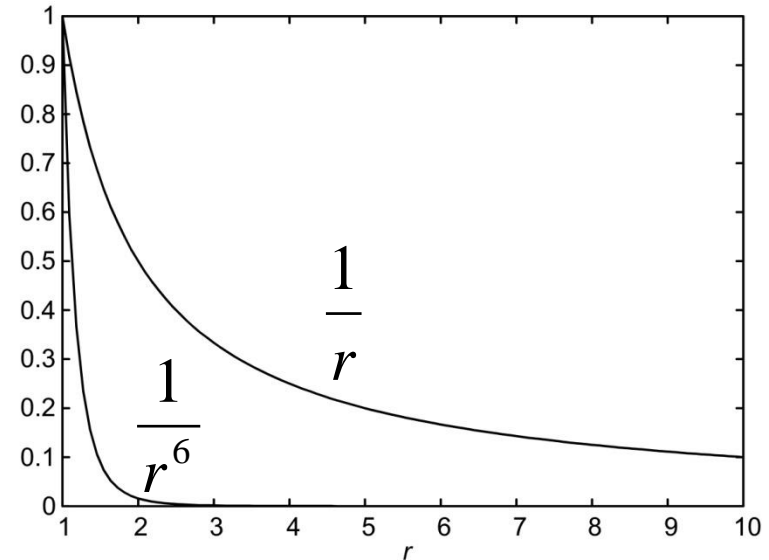
Minimum image convention



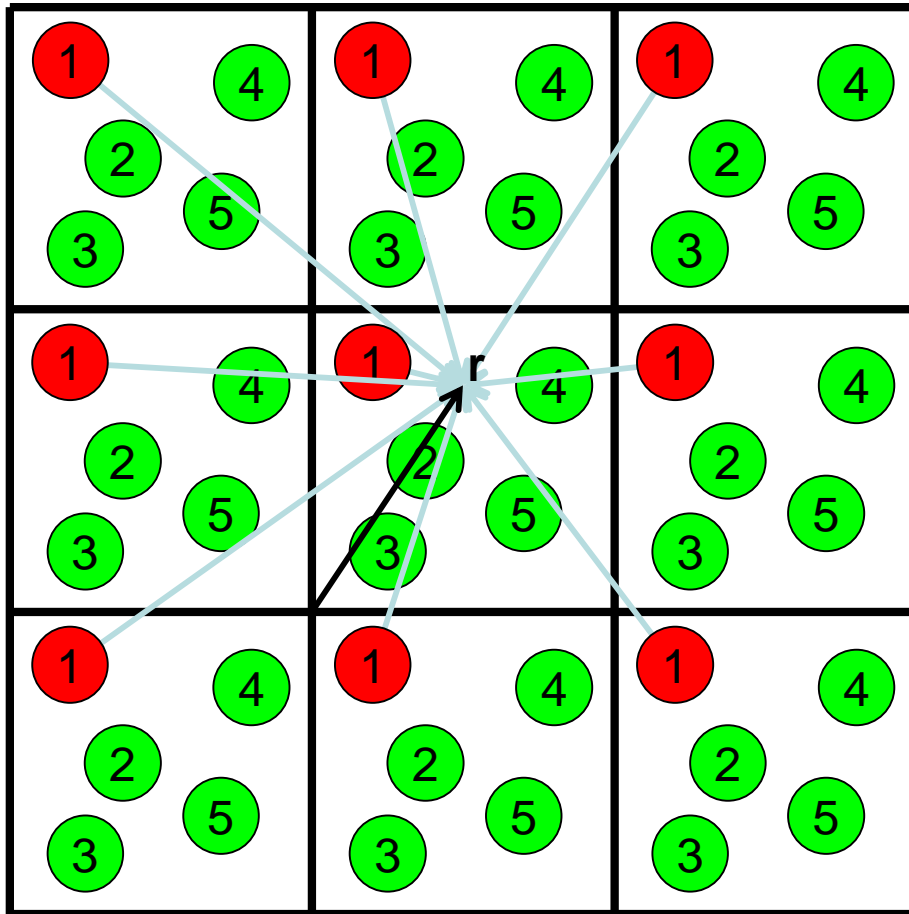
- カットオフ半径 r_c を最も短い基本セルの1辺の長さの2分の1以下にすれば考慮すべきペア数は $N(N-1)/2$ であり
→ minimum image convention

カットオフ法の問題点

- Van der Waals相互作用は遠距離では、 r^6 の項が支配的
→ van der Waals相互作用はカットオフ法で十分な精度で計算可能
- 静電相互作用は r^{-1} に依存
→ カットオフ法では精度良く評価することが困難
- 原子がカットオフ半径の範囲から出入りする際にエネルギーが変動するため、全エネルギーは保存しない



カットオフしない計算法(1)



中央の基本セル内の原子同士だけでなく、基本セル内の原子と周囲のイメージセル内の原子との間の相互作用も計算する

位置 \mathbf{r} における静電ポテンシャル:

$$\varphi(\mathbf{r}) = \sum_{\mathbf{n}} \sum_j \frac{q_j}{|\mathbf{r} - \mathbf{r}_j + \mathbf{L}\mathbf{n}|}$$

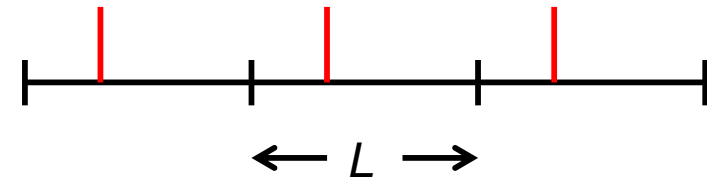
カットオフしない計算法(2)

- 周期境界条件では電荷は周期的に分布
 - 電荷分布をフーリエ級数で表すことができる

$$\rho(x) = \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left(2\pi i n \frac{x}{L}\right)$$

$$\tilde{\rho}_n = \frac{1}{L} \int_0^L \rho(x) \exp\left(-2\pi i n \frac{x}{L}\right) dx$$

$$\begin{aligned} \rho(x+L) &= \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left[2\pi i n \frac{(x+L)}{L}\right] \\ &= \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left(2\pi i n \frac{x}{L}\right) = \rho(x) \end{aligned}$$



カットオフしない計算法(3)

- 3次元の場合

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{\mathbf{n}} \tilde{\rho}_{\mathbf{n}} \exp\left(2\pi i n_x \frac{x}{L_x}\right) \exp\left(2\pi i n_y \frac{y}{L_y}\right) \exp\left(2\pi i n_z \frac{z}{L_z}\right) \\ &= \sum_{\mathbf{n}} \tilde{\rho}_{\mathbf{n}} \exp(2\pi i \mathbf{L}^{-1} \mathbf{n} \cdot \mathbf{r}), \quad \mathbf{L} = \begin{pmatrix} L_x & 0 & 0 \\ 0 & L_y & 0 \\ 0 & 0 & L_z \end{pmatrix}\end{aligned}$$

$$\tilde{\rho}_{\mathbf{n}} = \frac{1}{|\mathbf{L}|} \int_{\text{cell}} \rho(\mathbf{r}) \exp(-2\pi i \mathbf{L}^{-1} \mathbf{n} \cdot \mathbf{r}) d\mathbf{r}, \quad |\mathbf{L}| = L_x L_y L_z$$

カットオフしない計算法(4)

- Poisson方程式を解く

$$\nabla^2 \varphi(\mathbf{r}) = -4\pi \rho(\mathbf{r})$$

$$\tilde{\rho}_{\mathbf{n}} = -\frac{1}{4\pi|\mathbf{L}|} \int_{\text{cell}} \nabla^2 \varphi(\mathbf{r}) \exp(-2\pi i \mathbf{L}^{-1} \mathbf{n} \cdot \mathbf{r}) d\mathbf{r}$$

$$= \frac{\pi |\mathbf{L}^{-1} \mathbf{n}|^2}{|\mathbf{L}|} \int_{\text{cell}} \varphi(\mathbf{r}) \exp(-2\pi i \mathbf{L}^{-1} \mathbf{n} \cdot \mathbf{r}) d\mathbf{r}$$

$$= \pi |\mathbf{L}^{-1} \mathbf{n}|^2 \tilde{\varphi}_{\mathbf{n}}$$

$$\varphi(\mathbf{r}) = \sum_{\mathbf{n}} \frac{\tilde{\rho}_{\mathbf{n}}}{\pi |\mathbf{L}^{-1} \mathbf{n}|^2} \exp(2\pi i \mathbf{L}^{-1} \mathbf{n} \cdot \mathbf{r}) \quad \rightarrow \text{ポテンシャルエネルギー関数}$$

発散を防ぐために全電荷($\tilde{\rho}_0$)は0でなければならない 42

Particle Mesh Ewald法(1)

- n_x 、 n_y 、 n_z の範囲はマイナス無限大から無限大
- \mathbf{r} を離散化することで、この範囲を限定できる

$$\begin{aligned}\rho_{\mathbf{k}} &= \rho \left(\frac{k_x}{K_x} L_x, \frac{k_y}{K_y} L_y, \frac{k_z}{K_z} L_z \right) \\ &= \sum_{n_x=0}^{K_x-1} \sum_{n_y=0}^{K_y-1} \sum_{n_z=0}^{K_z-1} \tilde{\rho}_{\mathbf{n}} \exp(2\pi i \mathbf{K}^{-1} \mathbf{n} \cdot \mathbf{k})\end{aligned}$$

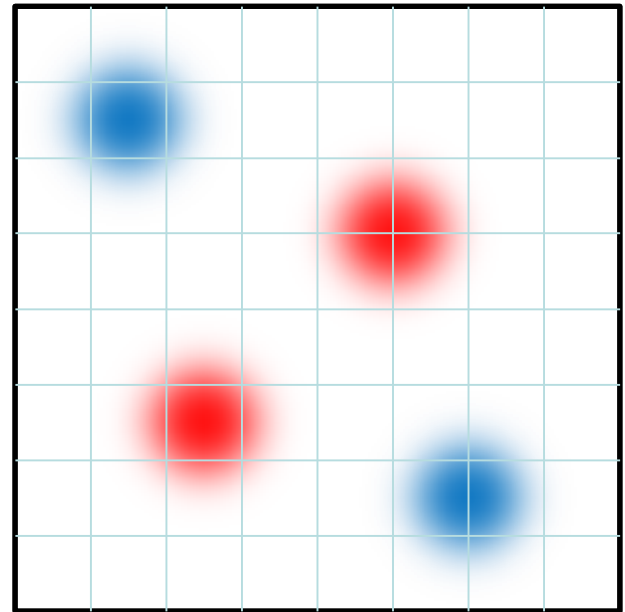
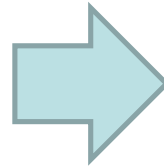
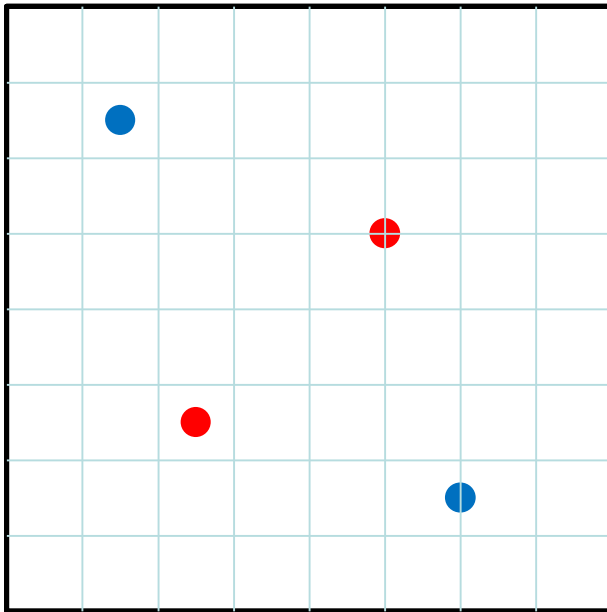
$$\tilde{\rho}_{\mathbf{n}} = \frac{1}{|\mathbf{K}|} \sum_{k=0}^{K-1} \rho_{\mathbf{k}} \exp(-2\pi i \mathbf{K}^{-1} \mathbf{n} \cdot \mathbf{k})$$

$$\varphi_{\mathbf{k}} = \sum_{n_x=0}^{K_x-1} \sum_{n_y=0}^{K_y-1} \sum_{n_z=0}^{K_z-1} \frac{\tilde{\rho}_{\mathbf{n}}}{\pi |\mathbf{L}^{-1} \mathbf{n}|^2} \exp(2\pi i \mathbf{K}^{-1} \mathbf{n} \cdot \mathbf{k})$$

} 高速フーリエ変換
で計算できる

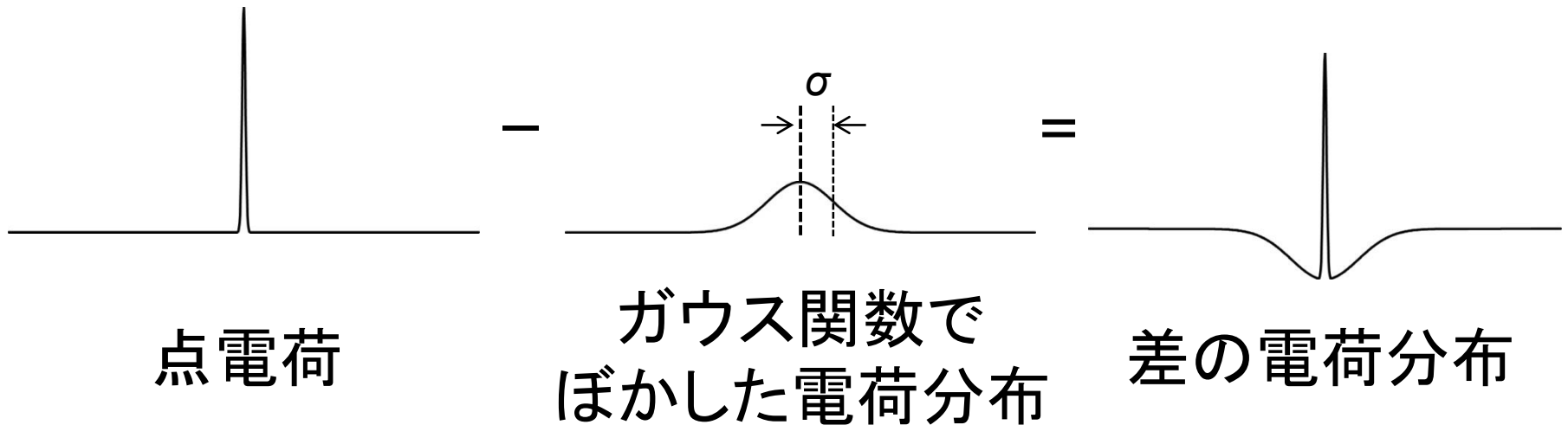
Particle Mesh Ewald法(2)

- 電荷分布は点電荷からなる
→ 一般には格子点に乗らない
→ ガウス関数でぼかす



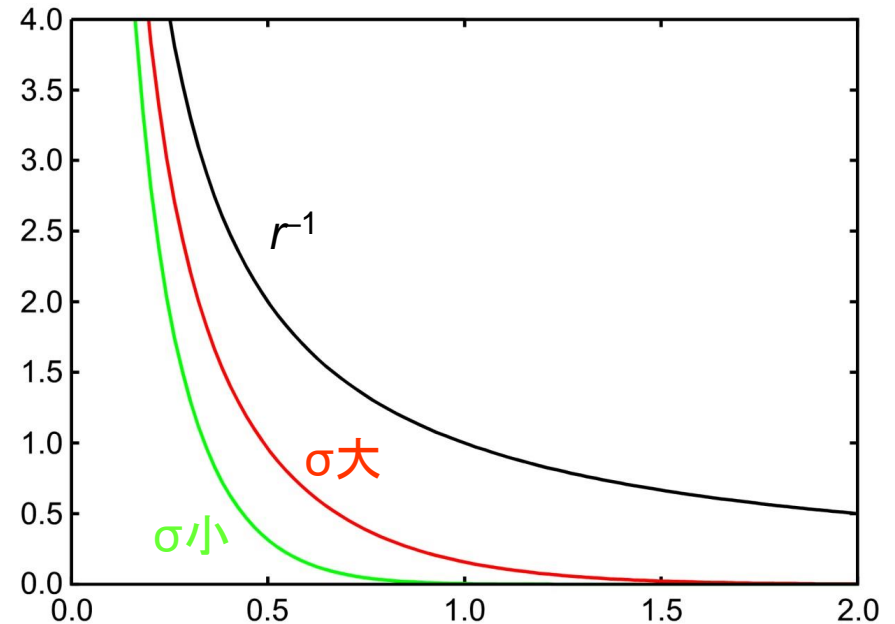
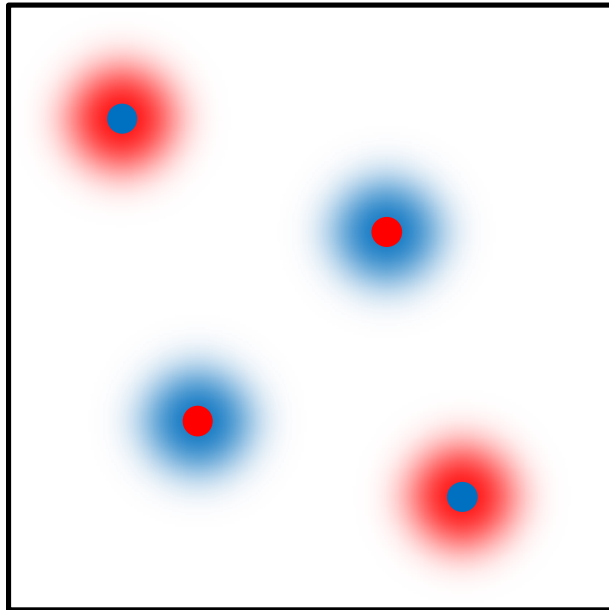
Particle Mesh Ewald法 (3)

- 本来の電荷分布からの差を求める



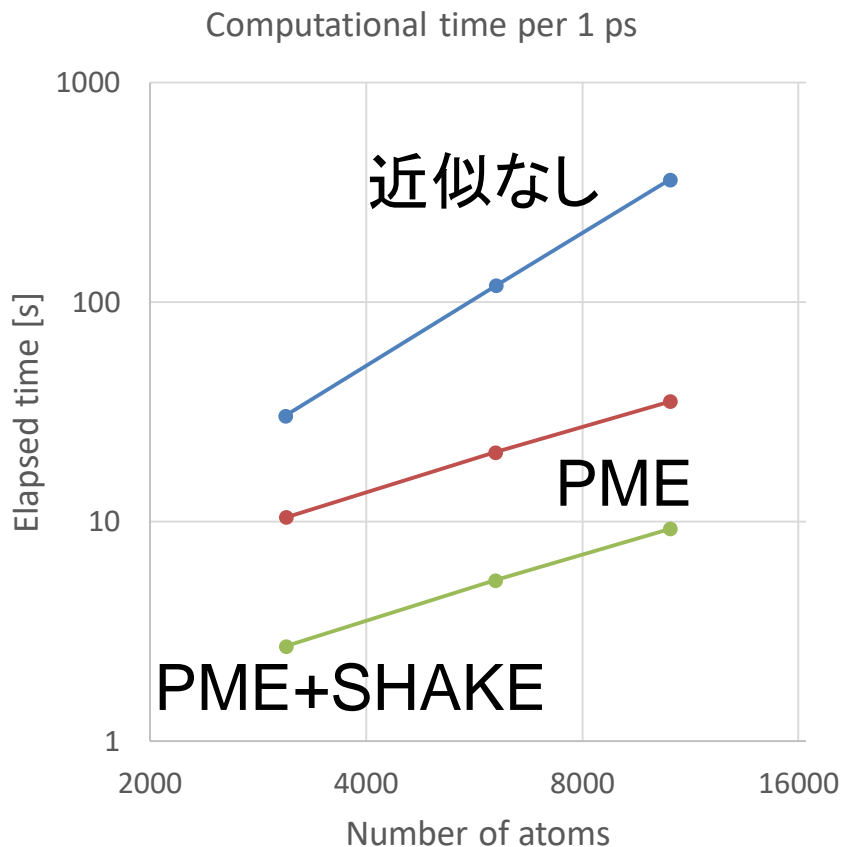
- 静電ポテンシャルを、ガウス関数でぼかした電荷分布がつくる静電ポテンシャルと、差の電荷分布がつくる静電ポテンシャルの和で表す

Particle Mesh Ewald法(4)



- 差の電荷分布では、点電荷のまわりに、これを打ち消す反対の符号の電荷が分布
 - 静電ポテンシャルは r^{-1} より速く 0 に減衰
 - カットオフ法でも精度よく計算できる

計算時間(2)



- 水分子の系で計算時間を計測
- 「近似なし」では原子数 N の2乗に比例
- PMEを使用することでほぼ $M \log N$ に比例
- SHAKEを併用することで時間刻みを4倍(2 fs)にでき、計算速度は3.8倍程度高速化した

NAMDにおける設定(2)

- PME法を使う場合は以下の設定を行う

```
cutoff                10.0
switching             off
cellBasisVector1     42.3810  0.0  0.0
cellBasisVector2     0.0  36.4706  0.0
cellBasisVector3     0.0  0.0  42.1148
PME                   yes
PMEGridSizeX         45
PMEGridSizeY         40
PMEGridSizeZ         45
extendedSystem        XSC_file_name
```

}★

☆

- ★と☆はいずれかを記載

タンパク質のMDシミュレーション(1)

1. ChimeraでPDB ID 1CRNの構造を開く
2. Stick表示に変更する
3. 水素原子を付加する
4. 水分子を直方体状に配置する
5. 電荷を付加する(標準残基の力場パラメータにAMBER ff14SBを指定)
6. パラメータファイルを保存する(ファイル名は1CRN、力場パラメータはAMBER ff14SB)

タンパク質のMDシミュレーション(2)

7. PDBファイルを保存(ファイル名: 1CRN.pdb)
8. 講義のページから1CRN.zipをダウンロードし、デスクトップに解凍
9. 生成されたフォルダを開き、先程保存した1CRN.prmtop、1CRN.inpcrd、1CRN.pdbを移動
10. restraint.plを実行→1CRN_rest.pdbが生成
11. min1.inのセルのサイズを修正
12. run.batをダブルクリックし、実行(約6分)

課題2

- 初期構造からのCa原子のずれ(RMSD)の時間変化をプロットせよ
- 平衡化(eq1、eq2)とプロダクション(prod)における、温度(TEMP)とポテンシャルエネルギー(POTENTIAL)の時間変化をプロットせよ
 - 時間刻み Δt は2 fs
- これらのプロットから何が言えるか考察せよ

課題の提出

- 課題1、課題2の結果と考察を1つのPowerPointファイルにまとめて、寺田宛 tterada@iu.a.u-tokyo.ac.jp に送ること
(エクセルファイルはサイズが大きいため送らないこと)
- その際件名は「分子モデリング課題」とし、本文に氏名と学生証番号、受講生IDを明記すること